NASA
Technical
Paper
2838

1988

Spatial Vision Processes: From the Optical Image to the Symbolic Structures of Contour Information

Daniel J. Jobson
Langley Research Center
Hampton, Virginia

National Aeronautics
and Space Administration

**Scientific and Technical
Information Division**

Abstract

The significance of machine and natural vision is discussed together with the need for a general approach to image acquisition and processing aimed at recognition. An exploratory scheme is proposed which encompasses the definition of spatial primitives, intrinsic image properties and sampling, two-dimensional edge detection at the smallest scale, construction of spatial primitives from edges, and isolation of contour information from textural information. Concepts drawn from or suggested by natural vision at both the perceptual and the physiological level are relied upon heavily to guide the development of the overall scheme. The scheme is intended to provide a larger context in which to place the emerging technology of detector-array focal-plane processors. The approach differs from many recent efforts at edge detection and image coding by emphasizing edge detection at the smallest scale as a foundation for multiscale symbolic processing while diminishing somewhat the importance of image convolutions with multiscale edge operators. Cursory treatments of information theory illustrate that the direct application of this theory to structural information in images could not be realized.

Background

Much of the meaning we acquire from the world around us is derived from the sense of vision. Likewise, for the bulk of the animal kingdom, vision is so central to basic activities that survival in the wild crucially depends on the visual functions of organisms. Another measure of the pivotal role of vision is the apparent dominance of our most intricate and vital organ, the brain, by vision processes (ref. 1, an excellent introduction to natural vision). At an entirely different level, words such as "vision" and "image" have taken on meanings far beyond our immediate visual perceptions and are broad metaphors for qualities such as wisdom, overriding impression, style, essence, remarkable intellectual ability, extraordinary artistic talent, and the ability to anticipate the consequences of major decisions. In a like manner, terms such as "insight," "picture," and "far-sighted" occupy a territory in our language reserved for our highest or most sweeping judgments. The sophistication of the visual tasks we routinely perform is effectively masked by the smoothness and ease with which they are carried out. In contrast the technology of machine vision is remarkably primitive in capability. A machine capable of one sophisticated visual task, such as reading or arbitrary object recognition, would be considered a technological marvel and yet remains beyond our grasp. Major advances in machine vision have far-reaching applications, with a broad, pervasive impact on economic development, especially in automation. The current state of machine-vision technology is characterized as a highly developed general capability for acquisition, transmission, and storage of images through the use of electronic media and a few specialized capabilities for performing visual tasks primarily in extremely cooperative or restricted situations.

Much progress has been made in advancing scientific knowledge about natural vision; however, a comprehensive, conclusive understanding of vision processes related to anatomical structure, physiological functions, and visual perception has been as elusive as major advances in machine-vision technology. The pursuit of this knowledge must be regarded as a premier scientific goal for our time because successful efforts would, in large measure, answer central questions about the brain.

The relationship between natural and machine vision is not a direct one and deserves further discussion. Both fields share the same basic questions, which are (1) what information must be isolated from the optical image to form the basis for performing visual tasks, and (2) how is this information extracted from the image? Another common question, not directly associated with the optical image, is how is visual memory organized and the act of comparison or interpretation performed? Ultimately no complete separation of natural vision and machine vision is either possible or desirable. Since no general theory of visual information has been accepted, machine vision must either directly or indirectly refer to natural vision. Usually the machine-vision references to natural vision are at the level of visual perception and invoke assumptions or definitions. Assumptions such as "edges in images are important" are derived from compelling visual perceptions rather than from physical laws. More recently some machine-vision research (ref. 2) has been stimulated by the physiological level of natural vision, that is, the organization of retinal and cortical receptive fields. The difficulty of using physiological data lies with its incompleteness and the fact that the machinery of natural vision is radically different from man-made cameras and computers. One difference, however, is highly encouraging—the natural-vision system is built from nerve cells with extremely slow signal responses and transmissions compared with those of the components of electronic circuitry. More discouraging differences are the remarkable capacities of neural circuitry for intricate connections, submicron structures, and elegant control of elaborate, high-volume parallel processes with multiple levels of feedback and adaptability. A more embracing relationship between these two fields may arise from basic research in visual information. Major new discoveries in natural vision such as the cortical visual subsystem of cytochrome oxidase regions (refs. 3, 4, and 5) provide much food for thought for machine-vision researchers. This subsystem permeates the primary visual pathway of the cortex, is the most metabolically active part of the pathway, and is found only in primates. While this subsystem's ability to process color has been established, its spatial functions are still undefined. In a different vein, the image processing tools of the machine-vision researcher can be very useful in testing scientific hypotheses concerning natural-vision processes.

Introduction

The current investigation attempts to define a general approach to spatial vision processes. In so doing it relies heavily on natural-vision concepts and addresses the two questions of what spatial information contained in the optical image is required for visual recognition of objects and how it is to be extracted from the image. Following a general discussion, tentative answers are given together with the promising results of image processing experiments. A key assumption suggested by natural-vision retinal and cortical functions is that a general set of processes exists which can be applied to all images and which produces the information necessary for object recognition without the need to resort to specialized processes for special tasks. The processing scheme must be regarded as preliminary and requires much additional refinement and demonstration. Convincing demonstration for quite diverse images is absolutely essential in view of the absence of basic physical laws or theory for either derivation of processes or confirmation of results. This investigation stops short of addressing any questions concerning visual memory and the actual process of comparison to effect recognition. It is assumed that the extraction of a *concise* set of spatial information from an image will naturally assist in the future development of a practical memory and recognition scheme.

General Discussion

If the broadest view of sensing and perception is considered, the crux of acquiring signals from the physical world and processing them is the transformation of raw signals obtained from some physical world measurement into the symbols of knowledge. In speech recognition (ref. 6) this overall process involves the transformation of measured acoustical signals into the phonemes of language. Victor Zue's efforts are highly instructive in several regards. His work represents a bottom-up approach which relates the distinguishing spectral-temporal characteristics of the acoustical signal to a well-defined symbolic vocabulary. Of special interest is the crucial need to account for known distortions and the absence of any references to higher language structure (i.e., grammar or word context). The analogy to vision, if appropriate, suggests that more consideration should be given to the characteristics of optical images, the smallest scale of processing (and necessarily the first stage), and the most immediate transformation possible into a symbolic domain. The problem with any direct analogy between speech and visual recognition is that the joint spectral and temporal signal characteristics of sound are replaced with the two-dimensional spatial and spatial frequency characteristics of images in vision. Further, no defined symbolic vocabulary exists for vision, and far more complexity and arbitrariness can be expected with visual recognition than with speech recognition. In short, the problem of visual recognition is far more poorly defined and certainly far more complex than speech recognition.

Before we approach the problem of extracting visual information for recognition, the larger scope of the following visual tasks and information is explored for perspective.

1. Image acquisition, transmission, and reconstruction for human observers
2. Image acquisition, image compression, transmission, and reconstruction for human observers
3. Scene representation using abstract symbols for human observers
4. Scene representation with abstract symbols for machine perception, recognition, and scene description

The lowest level visual task already accomplished with current technology is acquiring, transmitting, reconstructing, and storing images with machines. All interpretation is by human observers. The next level, which is highly experimental, is image coding—acquiring, *compressing*, transmitting, storing, and reconstructing images with machines. The end product is still presented to the human observer in a more or less close approximation to its original form. Motivation for data compression proceeds from the well-publicized bulkiness of images as packages of data and the bandwidth limitations of transmission links. Optical fiber transmission will greatly expand bandwidth limitations, but with the possibility of bottlenecks developing in the electro-optical conversion processes. At the next higher level of sophistication the visual task of scene representation or image rendering with abstract symbols is a dramatic leap beyond image reconstruction. The interpretation is still by human observer; however, the observer is no longer presented with a reconstruction of the original image, but rather a rendering perhaps much like that of an artist's illustration is produced. The highest level of visual task is for the total scheme to be accomplished by machine, that is, representation with the use of symbols and recognition, description, and interpretation. In this case the human observer might receive only a printed report on the interpretation of an image, that is, visual information cast into very brief language such as the following: "Image 542 contains one object-side view of an automobile of unknown make. No license plate is visible. Do you wish to have any further details?" Needless to say this last case is a highly imaginary one in terms of current technological capabilities. There is considerable overlap in these general visual tasks, especially in the highest three levels. The results of recent image coding research (ref. 7) suggest that all these tasks could be served by one general-purpose image acquisition and low-level processing front end, provided that the transformation from symbolic representation back to a reconstruction of an image is possible. In particular, research on second-generation image coding is exploring image operators which are quite similar to natural-vision image operators and to those used for symbolic representation in machine-vision research. The critical point is whether image renderings drawn from symbolic representations can be constructed so they are as convincing as the original images without being highly accurate as a pixel-by-pixel reconstruction of intensity values. In any case an immediate goal of exact image reconstruction must be dropped in order to address the larger questions, namely, what is the information needed from the image which can be the basis for performing visual tasks, and specifically what information is needed for recognition?

The Choice of Contour Information as a Central Focus

Reference 8 expresses the logic which, on the whole, is similar to that used herein. The argument is that the information in line drawings is the primary basis for recognition in visual perception because observers can correctly interpret line drawings of images without resorting to color, shading, texture, stereometric cues, or monocularly viewed three-dimensional scene presentation. A distinction is made herein which should be emphasized. The line drawing is an explicit image itself, but it represents a more abstract form—the spatial layout of contour information in the image. Further, it must be emphasized that contour information differs from the elaborate line drawing which often contains considerable surface texture and shading effects. Contour information refers only to the significant zonal boundaries of an image and is represented by the most simplified line drawings such as those in coloring books and graphic visual aids. Contour information is skeletal and not specifically concerned with surface characteristics other than their boundaries, including the boundary between differently textured surfaces. This investigation, while agreeing with the starting point of the Walters' treatment and the emphasis on general visual processes (ref. 8), differs from her work in a fundamental way—namely, Walters concentrates mainly on processing line drawing images, while the concern herein is transforming natural gray-scale images into contour information which is merely represented in an explicit display as a line drawing.

We seek to define contour information by analyzing the elements of simple line drawings in a very quantitative way. To do this a grid pattern is selected. For this investigation a square grid layout is chosen as most representative of image spaces encountered with electronic images. For natural vision a hexagonal layout is undoubtedly more appropriate in view of the retinal structuring of photoreceptors and neural circuitry. Since the line drawing itself is an explicit representation of an abstraction (i.e., contour information), we assume that the widths of the lines can be made vanishingly small and are irrelevant to our general definition. The actual scale of the grid elements is also not particularly important. There is a smallest scale definable for any image based on the optical blur together with the image sampling scheme, but we could arbitrarily make a scene larger than this smallest scale in an image. For this investigation, the smallest scale information in an image is defined as the information directly obtainable from one discrete image sample and its immediate neighbors, without interpolation. This consideration is of critical importance in the processing of an actual image but is not relevant to a general definition of information content other than to note the existence of a smallest scale limit. For a definition of the elements of contour information, arbitrariness of scale is a primary consideration. This does not mean that a coarse scale can provide as good a representation as a fine scale, but rather that the classes of elements themselves do not change with the scale. The following system of five elements (fig. 1) meets this requirement: null (N), simple line (S), shaped line (Sh), complex line (C), and end of line (E). Obviously, as we change scales for a particular line drawing, distributions of these elements change. As a general rule the shift from fine to coarse scales increases the relative proportion of complex line elements while it diminishes the proportion of null elements. Note that the choice of a hexagonal rather than a square layout does not affect the definition of this system of line elements.

The potential importance of this system is that an intrinsic logic of nearest neighbor groups can be established. For a nearest neighbor group of nine elements,

many permutations of elements are strictly forbidden and many others are rarely to be encountered. An example of a forbidden permutation is

N	N	N
S	S	N
N	N	N

The center S should have been an E. The total number of permutations of 5 classes taken 9 at a time is approximately 2.0 × 10^6 for a square grid. Interestingly, the hexagonal samples produce 5 classes taken 7 at a time, or approximately 7.8 × 10^4 permutations. This quantity of permutations is not nearly as discouraging as the quantity for the square grid, but it is hardly inspiring. Indeed the direct application of information theory must be delayed in the face of this seemingly intractable situation, especially since a common tool in information theory analyses is that equal probabilities of occurrence for all permutations can be assumed. Clearly this assumption of equal probabilities is not possible for this system of small groups of line elements, and the theoretical determination of probabilities of occurrence appears to be quite difficult, if not impossible.

The construction of definite rules governing classification of contour elements and allowable groupings of nearest neighbors will be considered after the gray-scale image and its edge events have been examined. The pathway from image acquisition through edge detection operations to contour extraction from spatial primitives is now considered.

Two-Dimensional Edge Detection and Representation

The questions which must be answered prior to performing edge detection on specific images are (1) what scale or scales should be chosen for edge operators, (2) what type of operator or operators should be chosen, (3) how is the operator to be constructed, given a sampled optical image, and (4) what technique should be employed for detecting and representing edges? Each of these questions is discussed in a logical sequence followed by the results of image processing experiments.

Scale

Although convolutions of multiple-scale edge or image-encoding operators with sampled optical images have enjoyed considerable popularity (refs. 2 and 7), the primacy of the smallest scale set by the spatial resolution limitations of the optical blur function and the image sampling scheme has not been sufficiently analyzed. The finest detail structure, along with much larger scale structure, is available only at the smallest scale, with the exception of certain image features or image conditions. These exceptions are (1) extended edges with a signal difference that is on a par with or below noise levels and (2) significantly blurred edges (i.e., certain shadows) or out-of-focus portions of a scene. These can and do occur in many images encountered, but these two phenomena rarely dominate the image unless there is a global defect in image quality. As a result the convolutions from larger scale edge operators are expected to be a necessary but often secondary engine in the machinery of vision. Edge detection at the smallest image scale is therefore explored, with a new emphasis, for capture of the finest detail and many larger scale features as well.

Choice of Edge Operator

A circular operator (fig. 2), which has been descriptively referred to as a Mexican hat function, is selected for uniform sensitivity to edges of arbitrary orientation, uniform suppression of two-dimensional low-spatial-frequency signals, its zero-crossing edge detection properties (ref. 2), and its ubiquitous occurrence in natural-vision retinal preprocessing (ref. 9). Various mathematical forms, which are essentially equivalent, have been used. Herein the mathematical form of a Gabor elementary signal is used since it places a precise form of the function in the framework of a full theory of communication (ref. 10). Further, the Gabor elementary signals have in various forms proven to be useful and accurate models for neurophysiological processing in natural vision (ref. 11) as well as in other sensory processes (ref. 10). A circular two-dimensional elementary signal takes the form of

$$G(r) = \exp(-r^2/2\sigma^2)\cos(2\pi f r) \qquad (1)$$

where $r = (x^2 + y^2)^{1/2}$, σ is the Gaussian space constant, and f is the modulation frequency. The terms

σ and f are reciprocally related for any particular form of $G(r)$ and the product of σf must be defined. If we invoke the constraint that the area integral over two-dimensional space must be zero to fully extinguish zero-spatial-frequency signals, a unique value of $\sigma f = 0.2080$ is found (fig. 3). The cross section of this elementary signal (fig. 4) is essentially the same as other mathematical functions often used (i.e., difference of Gaussians and Laplacian of Gaussian). This form establishes the exact reciprocal relationship between the space constant and the modulation frequency which must apply for any choice of spatial scale.

Construction of Edge Operator From Weighted-Image Samples

All scene radiance distributions have undergone two-dimensional convolution with the optical blur function and detector-array aperture response, and the sampling process has preset the amount of overlap between adjacent samples. A full mathematical treatment of the contributions of the optical blur function and the detector-array aperture functions to sampled images (ref. 12) contains an example of constructing a smallest scale difference of Gaussians (DOG) operator. A particular choice of weights for a 3×3 group of image samples with a specific amount of blur and square detector-array apertures achieves an excellent smallest scale DOG. The image samples processed with the weights are then equivalent to the sampling of a two-dimensional convolution of the DOG function with the scene radiance (intensity) field. Therefore, for one case the specific DOG (or, equivalently, the circular Gabor elementary signal) can clearly be constructed at the smallest scale in an image.

An attempt is now made to extend this special case toward more generality. How difficult is this construction for most digital images, where optical blur, detector-array or television vidicon point-spread function, and sampling lattice may all be unknown? Wide variations in detector element geometries are common, so a group of nine Gaussian functions in a square grid is examined as a representative case for imaging systems where the optical blur function dominates the detector aperture response in the overall two-dimensional character of the system response. The results (see fig. 5) are shown for a group wherein spacing is varied uniformly over a wide range relative to the optical blur space constant σ_B of the individual optical Gaussians. These results are also equivalent to maintaining a constant spacing and varying σ_B. Only half of the cross section of the resulting two-dimensional function is shown, and it exhibits excellent shape quality compared with the circular Gabor elementary signal (or, equivalently, the DOG). The constraint of a zero-areal integral is maintained, being less than 0.1 percent of the value for the individual Gaussian. Circular symmetry is well preserved except for the value of $x = 2\sigma_B$. Therefore, a wide range of practical image conditions can be covered by the choice of one set of weights, even when optical blur and sampling are somewhat variable.

The fairly general application of this set of weights has important implications for the design of general-purpose, front-end detector-array image-plane processor hardware. The weights used here were determined by dividing the circular Gabor function into nine squares and integrating and normalizing the center square to unity and the two pertinent adjacent squares relative to the center square. A quite different empirical approach (ref. 12) has determined the same values for the weights. The hexagonal array should present an easier problem since it possesses intrinsic circular symmetry, equal-valued perimeter weights, and densely packed circular detector apertures.

Two-Dimensional Edge Detection and Representation at the Smallest Scale

In a discrete sampled image in which the circular Gabor function is constructed at the smallest scale, a discrete two-dimensional convolution of two two-dimensional functions results. This process is, in effect, a stepped integration of the circular Gabor function and the scene intensity or relative radiance distribution. Previous work (ref. 2) on detecting edges by zero crossings has emphasized larger scale operator sizes and more samples of edge convolution signals than are available at the smallest scale in discrete samples of images. As we continuously convolve an edge signal with the circular Gabor function along any direction other than parallel to the edge, a characteristic curve occurs (fig. 6). For the smallest scale discrete image samples, only a few points on this curve are available and their exact placement is completely arbitrary, but the relative spacing is determined by the image sampling lattice. The minimum number of samples available for each event is six in the local neighborhood of nine image samples (fig. 6(b)). It seems natural to question why six or more samples are needed to determine what seems to be three edge locations. We could take an essentially one-dimensional approach and sift through a line of convolution values and place an edge location wherever a zero crossing occurs. If we do this in the example, we are immediately faced with a dilemma. The zero crossing often occurs between two samples, so where do we put the edge location? Well, we could establish a convention and place all edge locations either to the right or to the left of the actual interpixel zero crossing. Consider another example—perhaps the simplest test of spatial resolution—two bars at the scale of the smallest image samples (fig. 7). If we apply this same treatment to the discrete convolution samples, the result is a perfectly meaningless dense and structureless mass of detected edge events! This example is particularly instructive since we detect all the edges correctly but fail to detect and represent something equally important—namely, where edges are *not located!*

Now if we return to the samples of the characteristic convolution edge signal in two-dimensions (fig. 6), we can find two clues to resolving the dilemma: (1) the peaks and valleys in the curve bear information about null locations adjacent to edge locations, and (2) a more two-dimensional approach to edge detection must be considered since each convolution sample is surrounded by adjacent samples offering many more possible comparisons. We must determine which comparisons are to be made and the character of the representation, both of which are necessary to preserve smallest scale resolved spatial structure. The representation must clearly provide for all edge locations and *all* adjacent null locations to retain unambiguous connectivity relationships. Such a representation is shown in figure 7 and requires a magnification factor of 2 over the original image sample space! We can now examine zero-crossing comparisons to find an approach which detects all edge locations and all adjacent null locations. This approach must consider that all possible edge and null locations include different locations within each image sample. The additional locations are not between pixels but rather reflect whether the edge falls more toward a specific sector of the periphery of the image sample as opposed to falling near the center of the sample.

A scheme for two-dimensional zero-crossing detection which is sufficient to produce this representation is illustrated in figure 8. This set of comparisons is made for each 3×3 group of image convolution samples by stepping one sample vertically or horizontally for each subsequent set of comparisons. Each comparison must include a test for opposing polarity and some definition of "zero." In the absence of other logic, the limits of values considered to be zero might best be set by the intrinsic noise level in an image which is either known in advance (where sensor performance is known) or estimated by examining areas of apparently uniform average value.

Edge Detection Image Processing Experiments

The approach and procedures for edge detection are applied to two images which are both quite different in pattern content from each other and which contain considerable diversity of pattern information within each image. In short, the two images represent wide-ranging arbitrariness of visual phenomena. The images were originally in color. The green band was selected as the gray-scale original image in each case. The toy-scene image (fig. 9(a)) contains lettering at about the scale of the individual pixel samples, clearly defined textural surfaces of various sorts in certain fabrics as well as mixed zones of contour and texture in the curtained backdrop, and object contours such as dolls and other toys, books, and illustrated objects on book covers. In contrast, the mandrill (fig. 9(b)) is a flurry of fur texture, some blurred and some in focus, with contour associated with facial features. The edge representations (figs. 10 and 11) for the two images derived from smallest scale circular Gabor convolution and two-dimensional zero-crossing edge detection are dramatic illustrations of the distance between the edge representation and an ultimate, concise contour representation.

A more finely detailed look at the edge representations shows a consistent trend of distortions. Many expected continuous edges are not detected perfectly as a result of insufficient signal change or insufficient sharpness. Likewise, the particular event associated with the intersection of two or more edges is often not detected perfectly. For this situation a gap is often produced in the zero-crossing maps, the gap apparently resulting from the higher contrast edge swamping the convolution at the intersection point. This gap reflects the inability to detect a form of singularity in the image distributions. A further distortion is the staircase appearance of oblique edges most noticeable in the book outlines of the toy scene. This appears to be an intrinsic defect of the square-grid sampling scheme, which presents a shaped artifact at the local neighborhood scale. These two distortions will have to be considered when edge event groups are classified into spatial primitives.

A further important point to note is the presence of many completely insignificant edge events (isolated small groups or individuals) or edge events of secondary interest (textural clouds, stipples, or hatchings). In summary this smallest scale edge detection process produces certain defects in contour information and a profusion of edge events not related to contour information; however, the edge representation is quite interpretable in spite of these defects. One defect which is not present is any significant spatial distortion of edge locations. This lack of distortion, together with the capture of fine detail, is the primary reason for using the smallest scale operator possible. These properties of the edge representation and the line elements defined from abstract line drawings can now guide the definition of generalized spatial primitives constructed from the edge representation as precursors to contour information.

Definition of Spatial Primitives Based Upon the Edge Patterns

The line elements derived from the highly abstract line drawings (fig. 1) are now referred to the edge events of gray-scale images. Further, the known distortions or limitations of the edge detection process must be treated. First, consider the abstract line drawing as an optical image with each line being about one pixel wide in the sampled images. The edge representation of this image and that of an equivalent gray-scale image is examined. For this case there is a major difference between the line elements of the abstract line drawing and of the edge representation—that is, each line is really two edges when resolved in the optical image. Spatial primitives must be revised based upon edges, not lines. The line elements are revised as follows and now refer equally well to edge patterns in both gray-scale images and line drawing images:

Element	Symbol
Null	N
Simple edge	S
Shaped edge	Sh
Complex edge	C
End of edge	E

Note that the original end-of-line element is now a special case of the shaped-edge element Sh. The end-of-edge element E is a dubious event since most images are of scenes with extended objects as their subject matter, and point phenomena (if resolved) would be a small circle or square with a null center. Therefore, E events are error situations in which some edge phenomena are not completely detected or resolved because of excessively crowded adjacent edge events, a shift to subthreshold contrast, or the presence of high levels of noise. As already noted for C events, actual image convolutions often produce a gap in the locus of zero crossings (hence, an E event) at the point of intersection. This does not mean that a C event could not occur in actual edge patterns, but rather that not all edge intersections produce them.

The detected 3 × 3 edge patterns of the two test images are now classified into the general spatial primitives. The results (figs. 12 and 13) illustrate the ability of this set of primitives to represent an extensive array of contour and texture phenomena even in regionally mixed groups. The color code is green for simple edges S, blue for shaped edges Sh, red for ends of edges E, yellow for complex edges C, and gray for null N. Shape artifacts of the square-grid matrix are *not* classified as shape primitives and the edge intersections which result in gaps (E events) are left as such. Blowups of portions of both images (fig. 14) illustrate the local structural consistency of the classification scheme. The spatial primitive distributions of each quadrant of both images (table 1) quantify an overall expected trend—the highest relative frequency of occurrence for nonnull events is for simple edges, which has a rather unexpected stability of 69 to 79 percent. The two "error" classes, E and C, vie with each other for the next highest relative frequency of occurrence, though this is not likely to be generally true for high-quality scenes with little or no textural phenomena. From the computational viewpoint, absolute frequencies of occurrence of nonnull spatial primitives are encouragingly low, ranging from 12 percent (toy scene) to 28 percent (mandrill) for the magnified representational space (1024 × 1024). Furthermore, if textural events are removed, we can expect these absolute values to drop precipitously for the mandrill and somewhat for the toy scene.

The application of edge detection and spatial primitive classification to images supplies the framework for a more detailed consideration of information theory. If we take only the nonnull center 3 × 3 edge patterns which are detected and their associated spatial primitives, a reasonably comprehensive analysis of permutations is now possible (fig. 15). Although absolute frequencies of occurrence cannot be assigned for the contour-texture representations, the symbolic vocabulary of possible 3 × 3 configurations can be narrowed dramatically with reasonable confidence. We can see a potentially more radical diminution of allowed choices for the spatial primitives than for the detected edge permutations. This diminution is due to the number of expected configurations without reference to their exact frequencies of occurrence. Determinations of frequencies of occurrence may not be possible in a general sense because they may vary widely from one image to another. This variability is shown in the contour-texture representations (figs. 12 and 13). A complete information theory analysis will only be possible if pure contour representations are achieved and display stable values for frequency of occurrence. This is most unlikely since pure contour itself can vary widely from sparse to dense patterns both regionally and globally. On the other hand, it may be possible to bracket frequencies of occurrence in some manner and thereby quantify information. This bracketing would require a more extensive analysis of diverse images and the distributions within their spatial primitive representations than is now being attempted. This endeavor is dropped in order to concentrate on methods for contour-texture discrimination that use the spatial primitive representation.

Contour and Texture Discrimination—An Issue for Multiscale Interpretation of Group Properties of Spatial Primitives

A completely satisfactory scientific definition of contour and texture does not exist; however, it is possible to establish the overall character of each and how they both can be formed from the same set of spatial primitives. Contour is most often composed of continuous connected simple S and shaped Sh primitives with rarely occurring complex C or end-of-edge E primitives. Contour information is often rather sparse, but it can be dense for cases such as printed text, where the line width of lettering is at or near the image sample size. Dense contour information such as text appears to have a complete absence of C primitives, a paucity of E primitives, but a significant number of Sh primitives for especially high densities. Texture likewise can be dense or sparse but is expected to possess a high percentage of C primitives for surfaces such as woven fabric or fur or a high percentage of E primitives for granular, ripped surfaces such as painted plaster or cinder block. Regular striped surfaces, grids, and gratings whose scale is small (near or below the image sample size) should also be considered as texture. Dense contour such as printed text becomes textural if it is made small enough to be poorly resolved. Likewise, grids and stripes become contour if made sufficiently large. We can cite examples of smallest scale 3 × 3 groups of spatial primitives which can be used to construct either contour or texture patterns equally well. Therefore, it is quite clear that scales larger than 3 × 3 must be examined to perform contour-texture discrimination.

Before engaging in an exercise in contour-texture discrimination, we can convey a more general feeling for the visual character of each. Obviously most texture examples cannot be complete or be expected to capture all types, but an attempt to compile typical or representative forms seems necessary. (See fig. 16.) Again, the distinguishing spatial primitive arrangements suggested are high frequencies of occurrence of E or C, or are just highly dense formations of S without much Sh occurrence. The visual perception of noise is textural, so noise is included in the texture classes.

A highly preliminary multiscale exercise in contour-texture discrimination is presented to illustrate the potential of this purely symbolic approach. (See figs. 17 and 18.) No attempt has been made to detect and represent the contours associated with purely texture boundaries. Therefore, these boundaries simply disappear. The methods used involve four scales—3 × 3, 6 × 6, 12 × 12, and 24 × 24 square windows—in the spatial primitive representations, and obviously further development is needed. These methods are based upon setting limits on the maximum total number of spatial primitives allowed for each scale and on the maximum number of E events allowed for each scale. However, most texture is dissolved while most contour is retained, so the promise of multiscale symbolic processing for contour-texture discrimination based on spatial primitive distributions is apparent.

The major limitation of the methods developed thus far is their inability to detect and represent all the contours that exist perceptually between differently textured zones and between a texture zone and adjacent null zones in images. This limitation is a subject for further investigation and it necessarily involves the question of large-scale windows. This investigation should relate any methods developed to the texton theory of Julesz (ref. 13), which has been used so successfully to treat the perceived differences between adjacent texture surfaces. The Julesz theory treats texture discrimination as being based on local densities and distributions of particular features in the image. In this sense textons are similar to the spatial primitives used herein; however, the orientation of edges figures prominently in the texton theory and is completely absent from this set of spatial primitives. Further, the Julesz textons refer to local features and structures of larger scale than the spatial primitives, which were developed as the smallest distinguishable units of structure. Both the textons and the spatial primitives used herein share an emphasis on spatial discontinuities (i.e., ends of lines and intersections for the textons and C and E events for the spatial primitives).

Conclusions

A general vision scheme encompassing spatial primitive definition, image acquisition, small-scale two-dimensional edge detection and representation, spatial primitive classification, and contour-texture discrimination was presented and illustrated with experimental image processing results. The scheme is intended as a preliminary set of methods for extracting from optical images the significant structural information that is necessary for subsequent machine object recognition and scene interpretation. Natural-vision concepts and visual perception were employed in the development of the scheme. A major limitation of the contour-texture discrimination approach is that no attempt is made to represent the contours which are perceived at texture boundaries. The principal results of this investigation are the following:

1. Both edges and adjacent null elements must be determined and require a representation space which is magnified by a factor of 2 over the original image space. Otherwise, the finest resolved structure in an image is lost.

2. General spatial primitives are defined which capture the smallest units of spatial structure in the optical image subject to limitations of the edge detection process in handling edge intersection discontinuities, artifacts due to a square-grid discrete image space, and ambiguities created by unresolved structure and noise.

3. The general spatial primitives can and do represent both contour and texture phenomena, and they improve the ability to discriminate one from the other through the use of multiscale symbolic processing.

NASA Langley Research Center
Hampton, VA 23665-5225
August 30, 1988

References

1. Frisby, John P.: *Seeing—Illusion, Brain and Mind.* Oxford Univ. Press, 1980.
2. Hildreth, Ellen C.: The Detection of Intensity Changes by Computer and Biological Vision Systems. *Comput. Vis., Graph., & Image Process.*, vol. 22, no. 1, Apr. 1983, pp. 1–27.
3. Livingstone, Margaret S.; and Hubel, David H.: Anatomy and Physiology of a Color System in the Primate Visual Cortex. *J. Neurosci.*, vol. 4, no. 1, Jan. 1984, pp. 309–356.
4. Horton, J. C.; and Hedley-Whyte, E. Tessa: Mapping of Cytochrome Oxidase Patches and Ocular Dominance Columns in Human Visual Cortex. *Philos. Trans. Royal Soc. London,* ser. B, vol. 304, no. 1119, 1984, pp. 255–272.
5. Tootell, Roger B. H.; Hamilton, Susan L.; and Silverman, Martin S.: Topography of Cytochrome Oxidase Activity in Owl Monkey Cortex. *J. Neurosci.*, vol. 5, no. 10, Oct. 1985, pp. 2786–2800.
6. Cole, Ronald A.; Rudnicky, Alexander I.; Zue, Victor W.; and Reddy, D. Raj: Speech as Patterns on Paper. *Perception and Production of Fluent Speech,* Ronald A. Cole, ed., Lawrence Erlbaum Assoc., Publ., 1980, pp. 3–50.
7. Kunt, Murat; Ikonomopoulous, Athanassios; and Kocher, Michel: Second-Generation Image-Coding Techniques. *Proc. IEEE,* vol. 73, no. 4, Apr. 1985, pp. 549–574.
8. Walters, Deborah: Selection of Image Primitives for General-Purpose Visual Processing. *Comput. Vis., Graph., & Image Process.,* vol. 37, no. 2, Feb. 1987, pp. 261–298.
9. Levick, W. R.: Receptive Fields of Retinal Ganglion Cells. *Physiology of Photoreceptor Organs,* M. G. F. Fuortes, ed., Springer-Verlag, 1972, pp. 531–566.
10. Gabor, D.: Theory of Communication. *J. Inst. Electr. Eng. (London),* vol. 93, pt. 3, 1946, pp. 429–457.
11. Daugman, J. G.: Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles. *Vis. Res.,* vol. 20, no. 10, 1980, pp. 847–856.
12. Huck, Friedrich O.; Fales, Carl L.; Halyo, Nesim; Samms, Richard W.; and Stacy, Kathryn: Image Gathering and Processing: Information and Fidelity. *J. Opt. Soc. America,* vol. 2, no. 10, Oct. 1985, pp. 1644–1666.
13. Julesz, B.; and Bergen, J. R.: Textons, the Fundamental Elements of Preattentive Vision and Perception of Textures. *Bell Syst. Tech. J.*, vol. 62, no. 6, July–Aug. 1983, pp. 1619–1645.

Table 1. Distributions of Detected Spatial Primitives

Event	Number of events (relative occurrence, percent[a]) for—				
	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4	Total image
			Toy scene		
S	27 986 (76)	25 691 (71)	16 628 (79)	22 169 (75)	92 474 (75)
Sh	1 491 (4)	1 458 (4)	645 (3)	1 211 (4)	4 805 (4)
E	4 458 (12)	7 851 (22)	2 945 (14)	4 165 (14)	19 419 (16)
C	2 664 (7)	1 360 (4)	743 (4)	2 039 (7)	6 806 (6)
All	36 599 (99)	36 360 (101)	20 961 (100)	29 584 (100)	123 504 (101)
			Mandrill		
S	60 339 (74)	57 091 (74)	46 421 (72)	48 006 (69)	211 857 (72)
Sh	5 010 (6)	4 740 (6)	3 625 (6)	4 663 (7)	18 038 (6)
E	5 664 (7)	5 774 (8)	9 333 (14)	8 450 (12)	29 221 (10)
C	10 442 (13)	9 053 (12)	5 011 (8)	8 657 (12)	33 163 (11)
All	81 455 (100)	76 658 (100)	64 390 (100)	69 776 (100)	292 279 (99)

[a] Accuracy of ±1 percent.

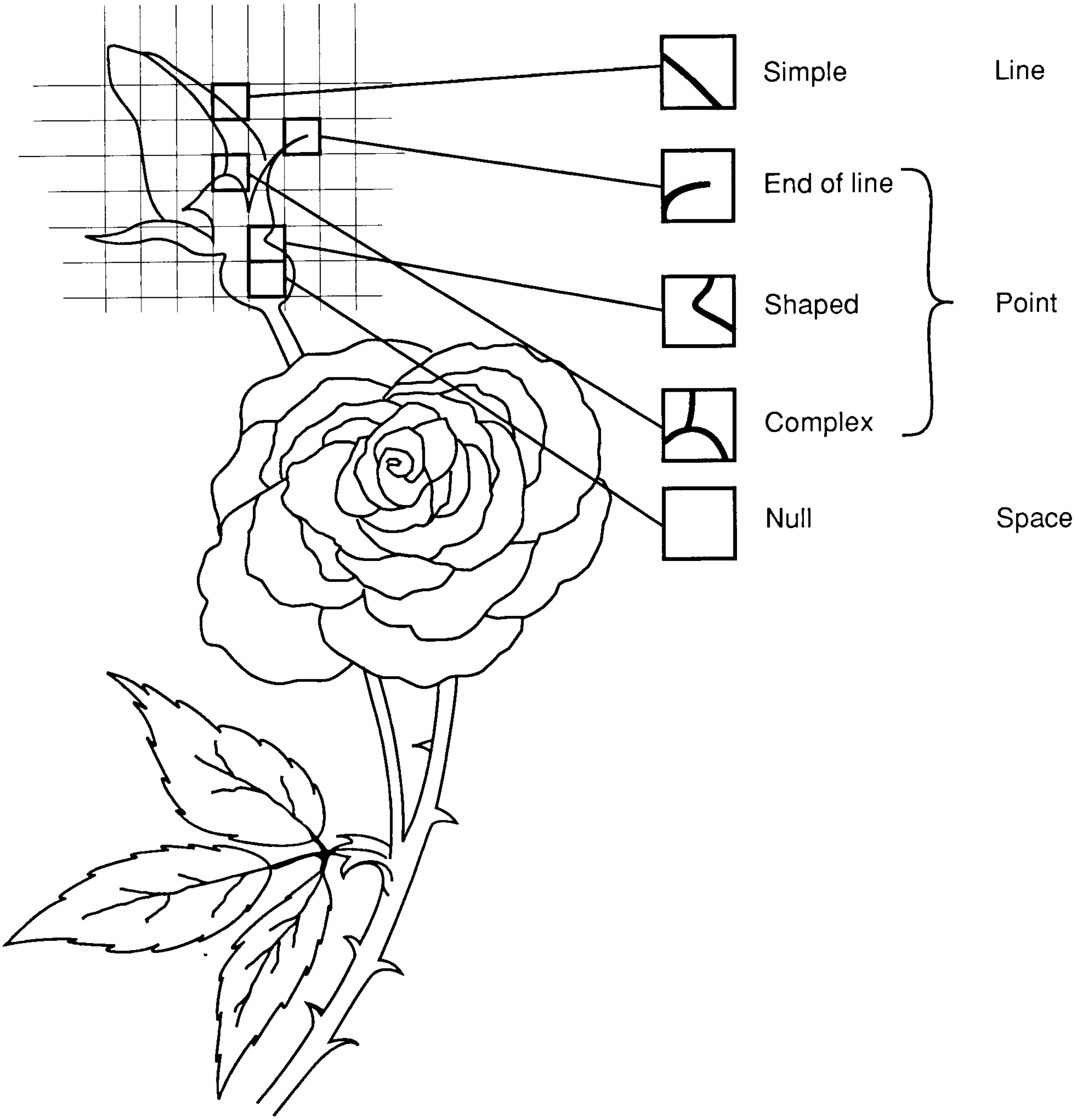
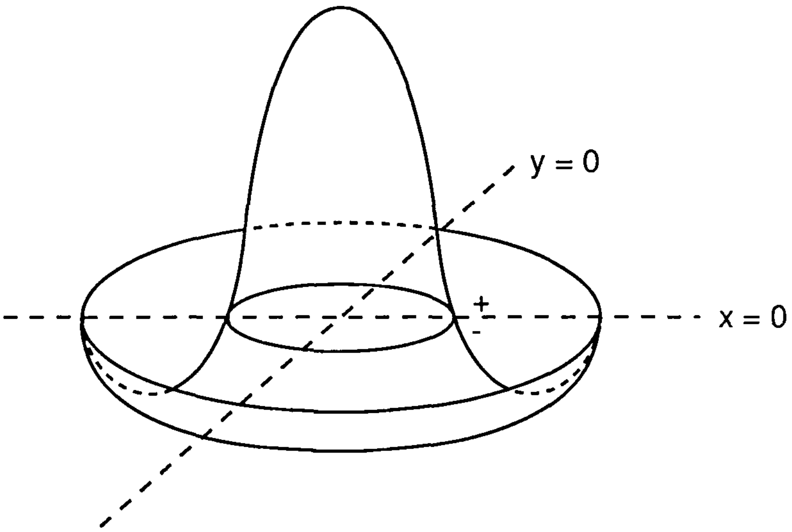
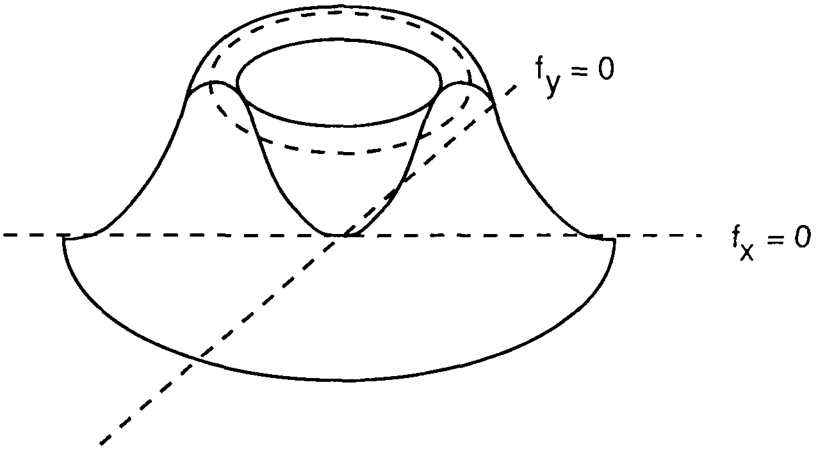


Figure 1. The elements of line drawing.



(a) Spatial domain.



(b) Spatial frequency domain.

Figure 2. Circular Gabor elementary signal desired for convolution with scene radiance.

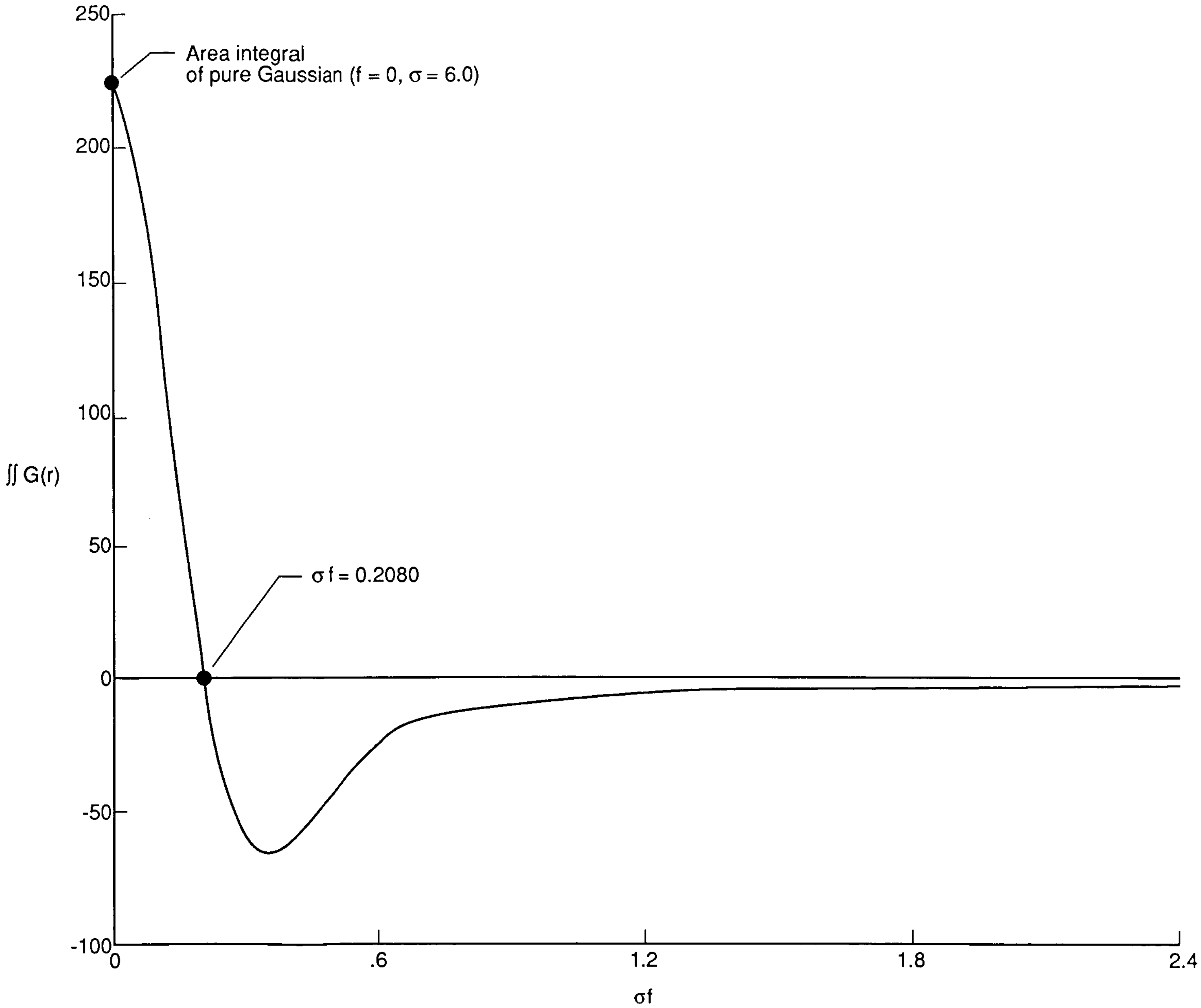


Figure 3. Area integral of $G(r)$ as function of σf.

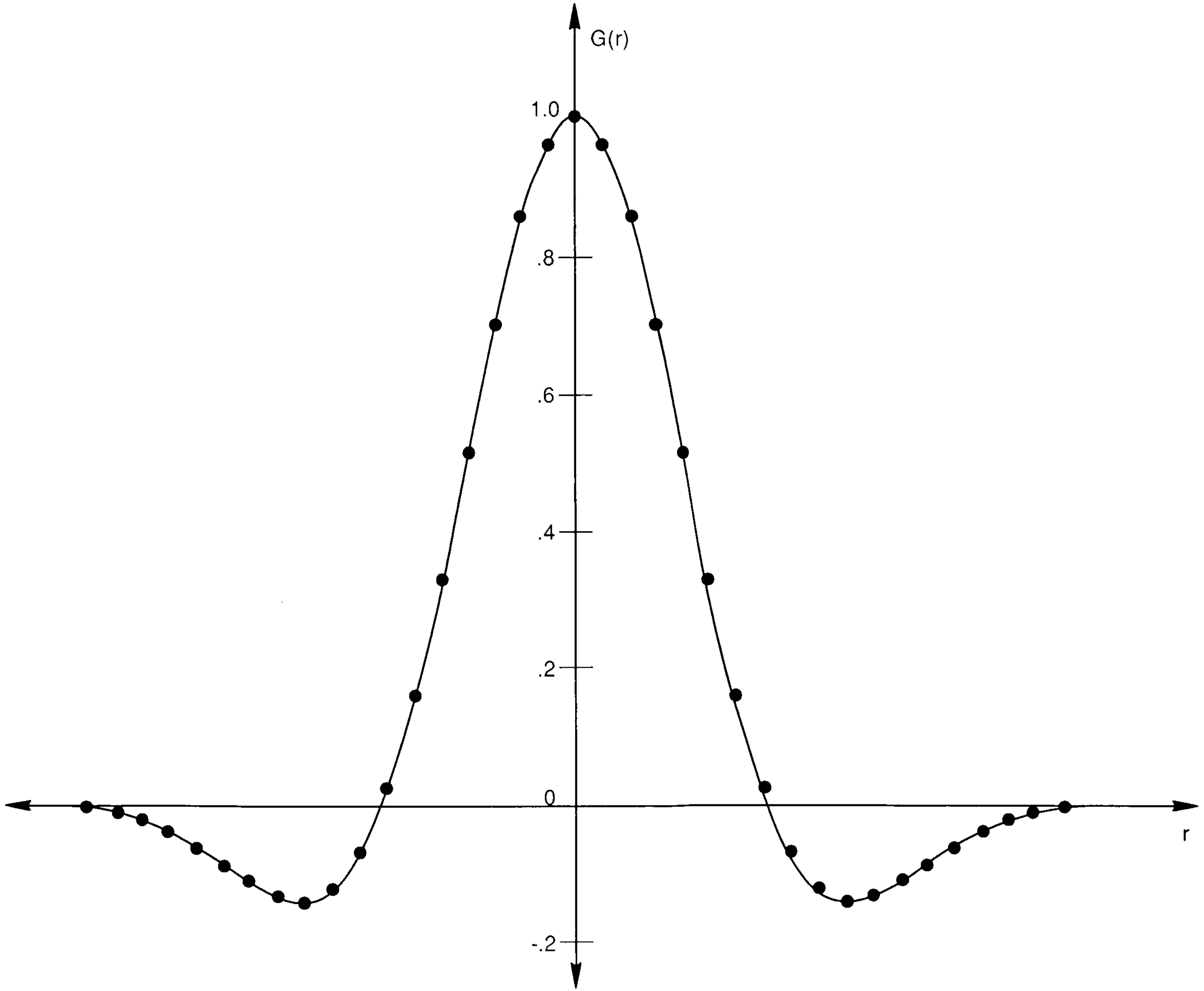
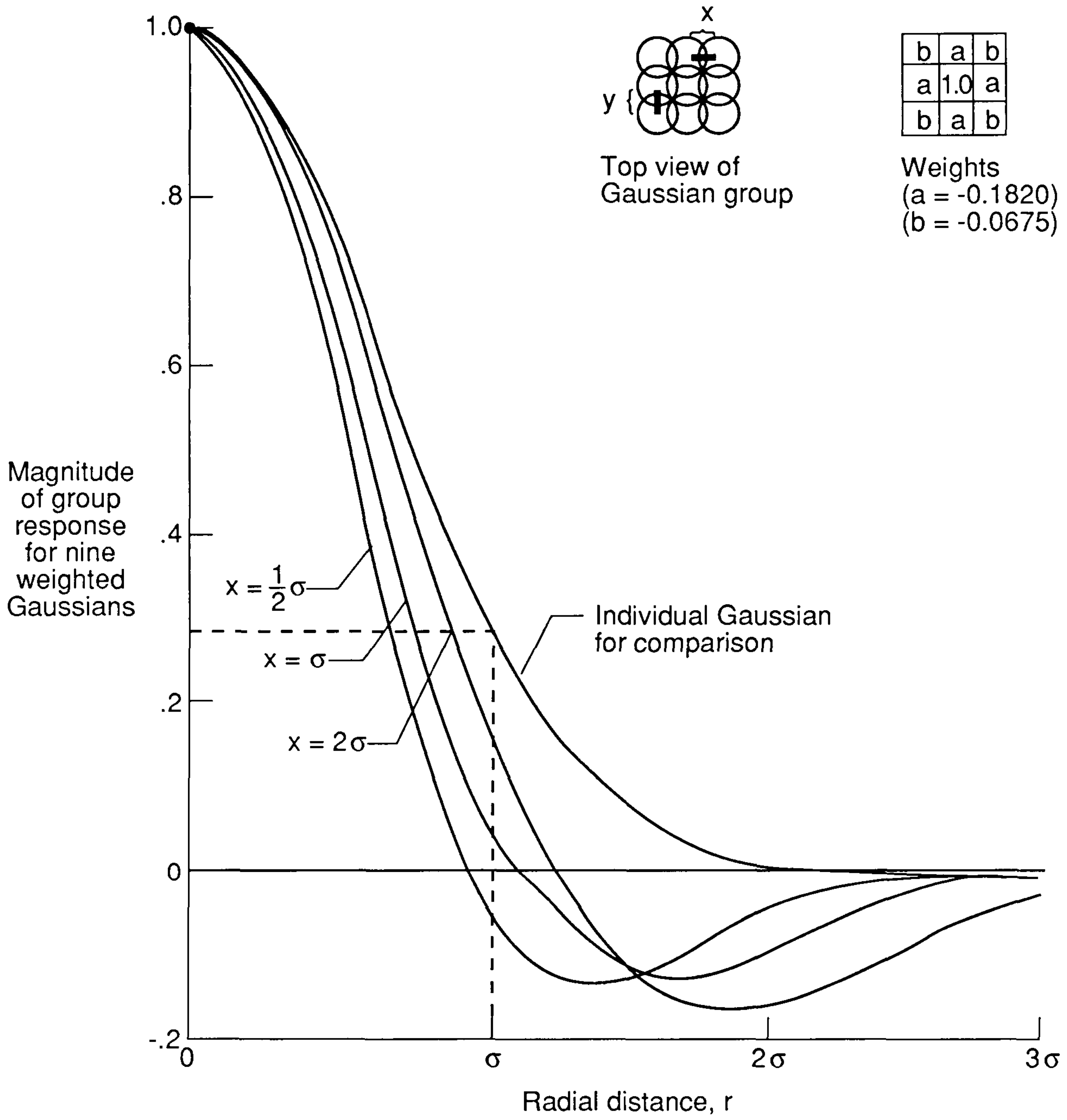
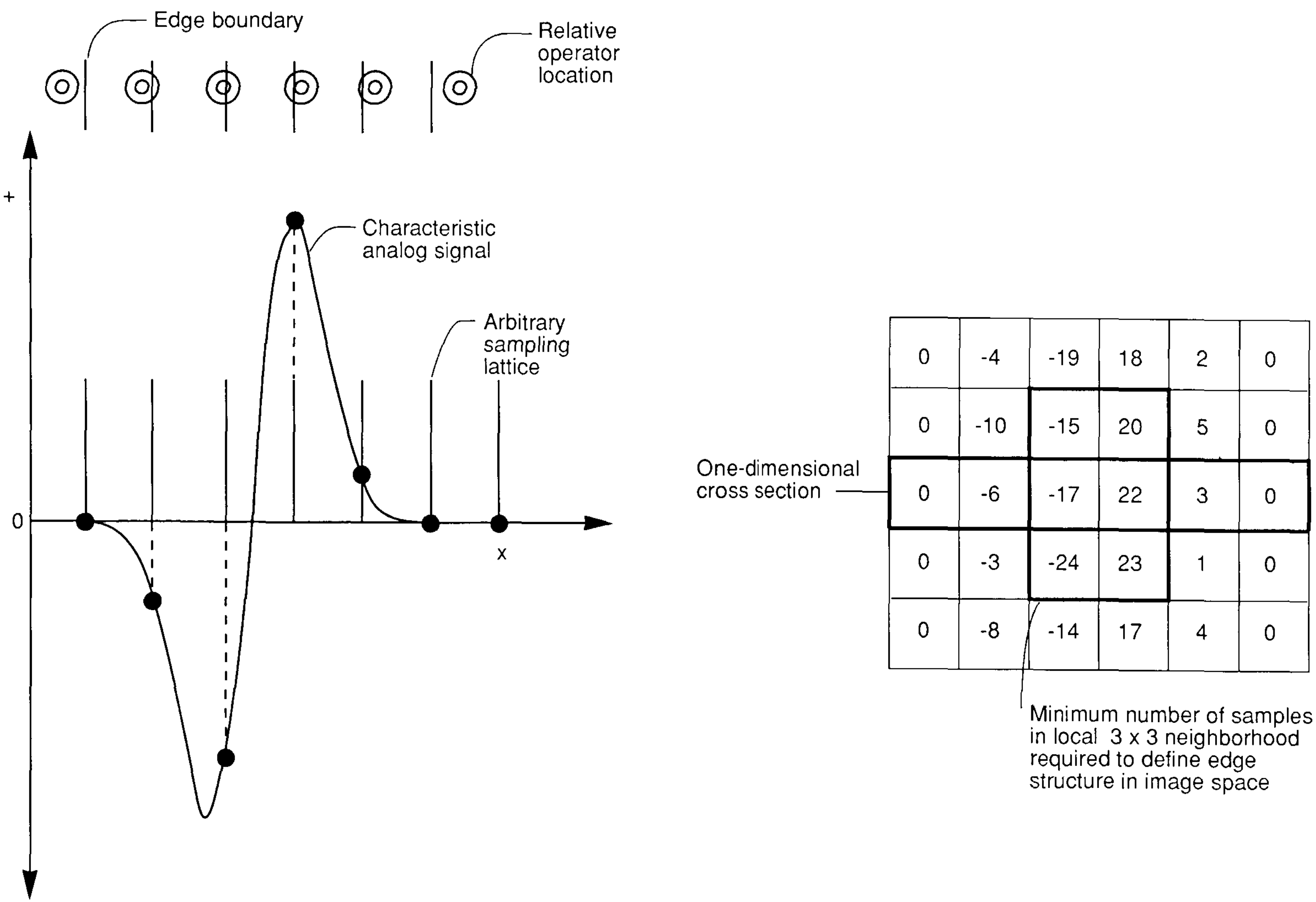


Figure 4. Cross section of two-dimensional circular Gabor elementary signal for $\sigma f = 0.2080$ with area integral value of zero.



b	a	b
a	1.0	a
b	a	b

Figure 5. Deformation of group response of nine weighted Gaussian functions for variable spatial overlap.

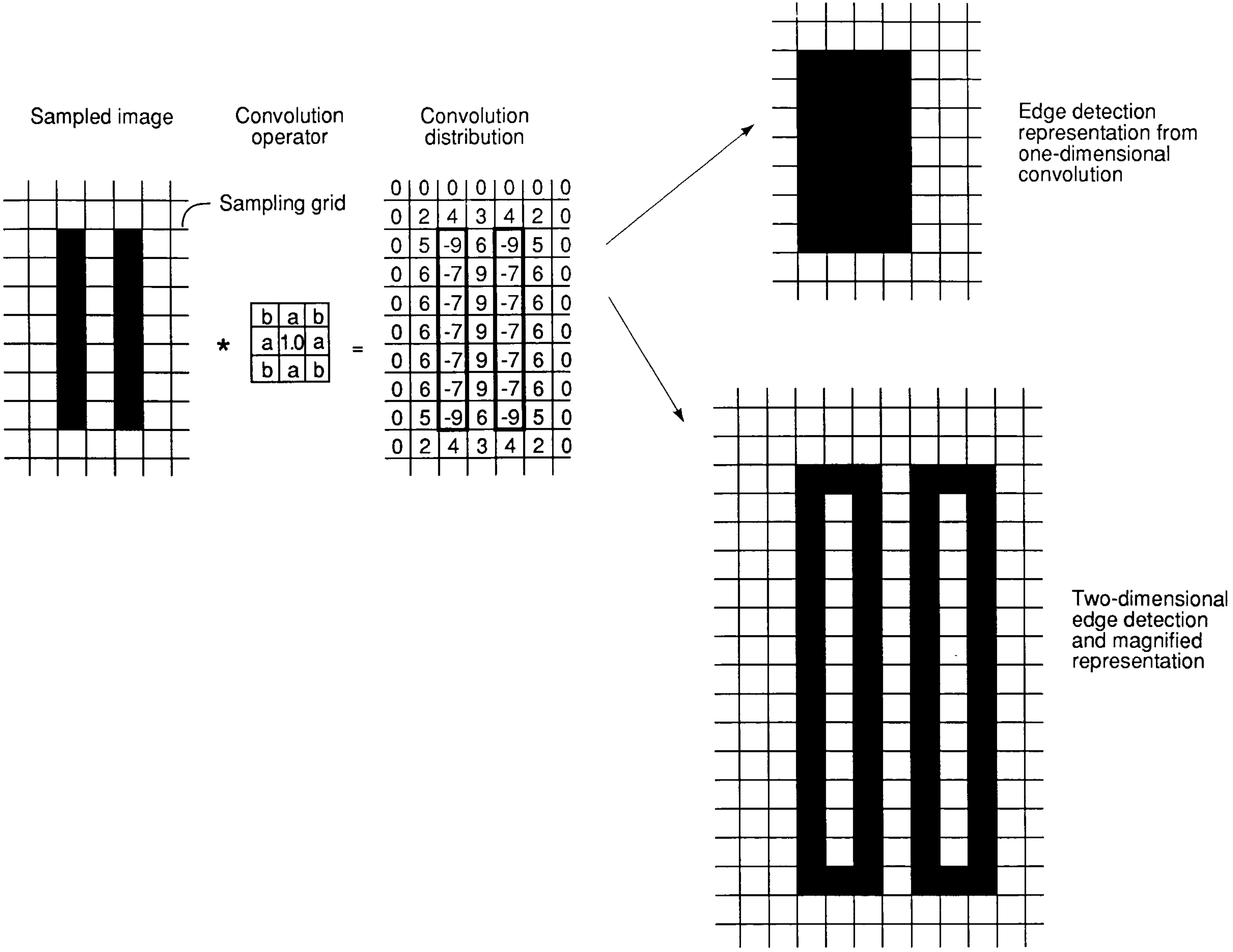

Edge boundary
Relative operator location
+
0
Characteristic analog signal
Arbitrary sampling lattice
x
One-dimensional cross section
Minimum number of samples in local 3 x 3 neighborhood required to define edge structure in image space

0	-4	-19	18	2	0
0	-10	-15	20	5	0
0	-6	-17	22	3	0
0	-3	-24	23	1	0
0	-8	-14	17	4	0

(a) One-dimensional cross section.

(b) Two-dimensional example of discrete samples.

Figure 6. Discrete sampling of convolution signals for scene edges.


Sampled image
Convolution operator
Convolution distribution
Sampling grid
*
=
Edge detection representation from one-dimensional convolution
Two-dimensional edge detection and magnified representation

b	a	b
a	1.0	a
b	a	b

0	0	0	0	0	0	0
0	2	4	3	4	2	0
0	5	-9	6	-9	5	0
0	6	-7	9	-7	6	0
0	6	-7	9	-7	6	0
0	6	-7	9	-7	6	0
0	6	-7	9	-7	6	0
0	6	-7	9	-7	6	0
0	5	-9	6	-9	5	0
0	2	4	3	4	2	0

Figure 7. Hypothetical comparison of one- and two-dimensional edge detection and representation processes for two-bar target.

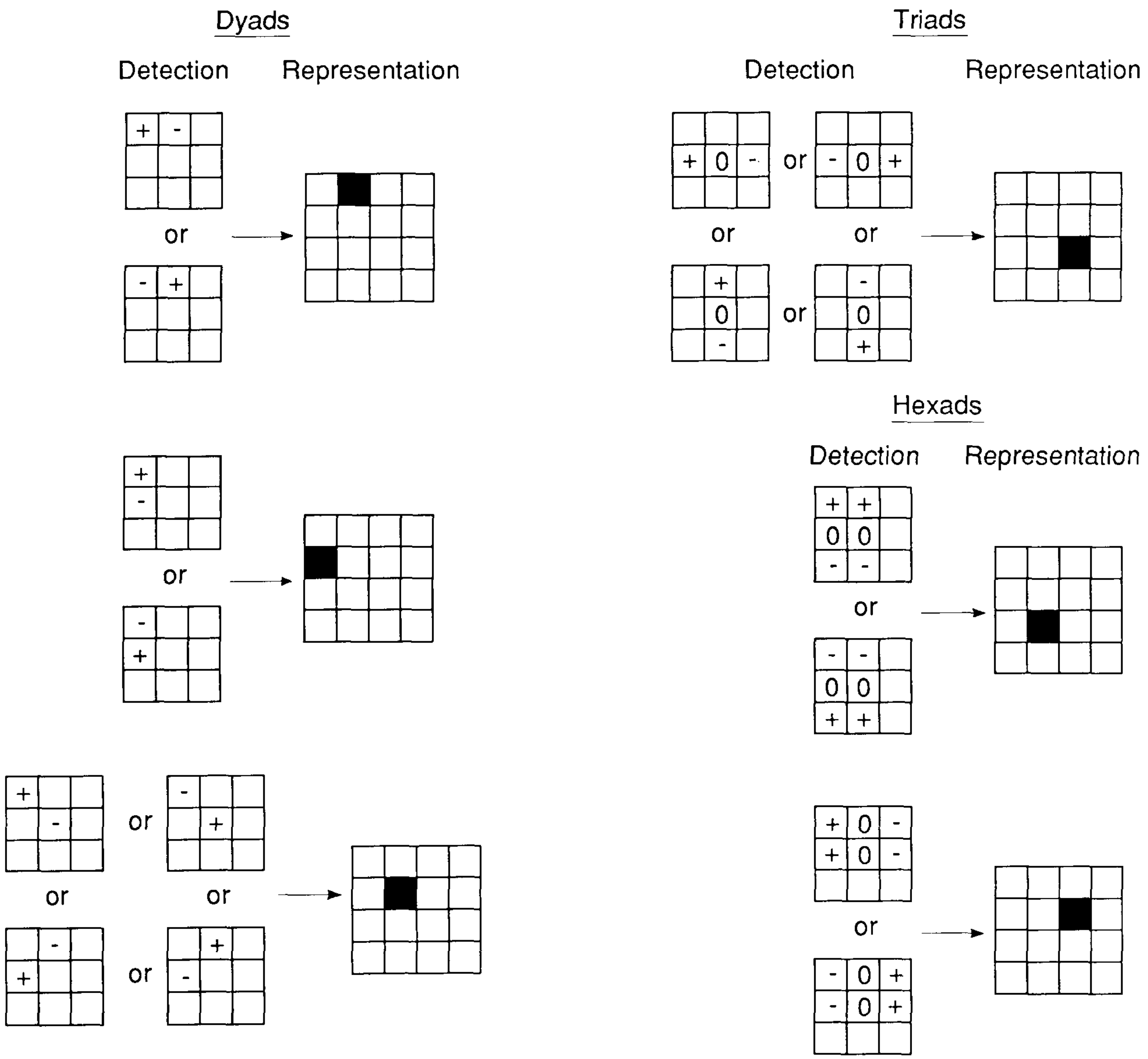
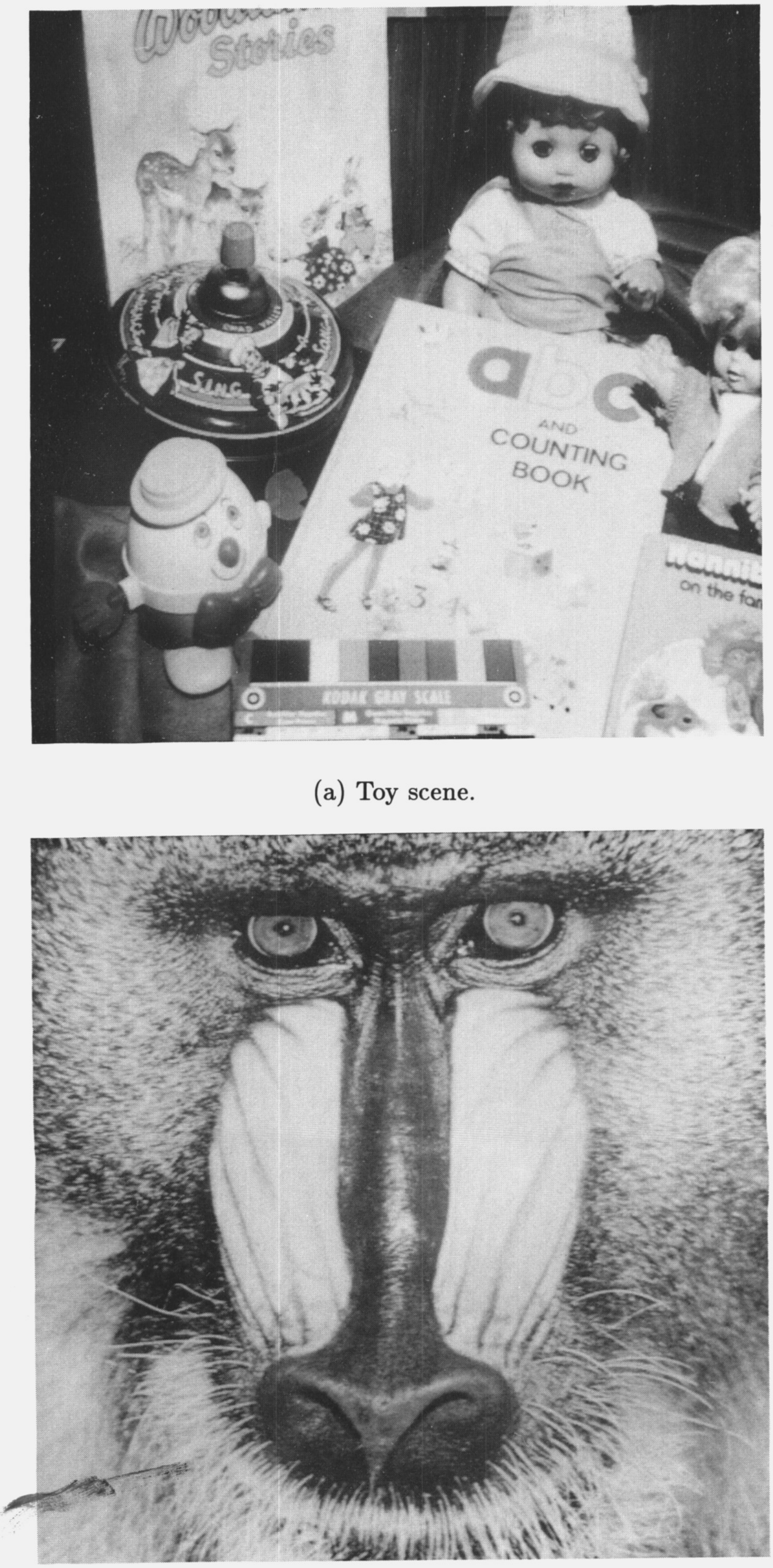


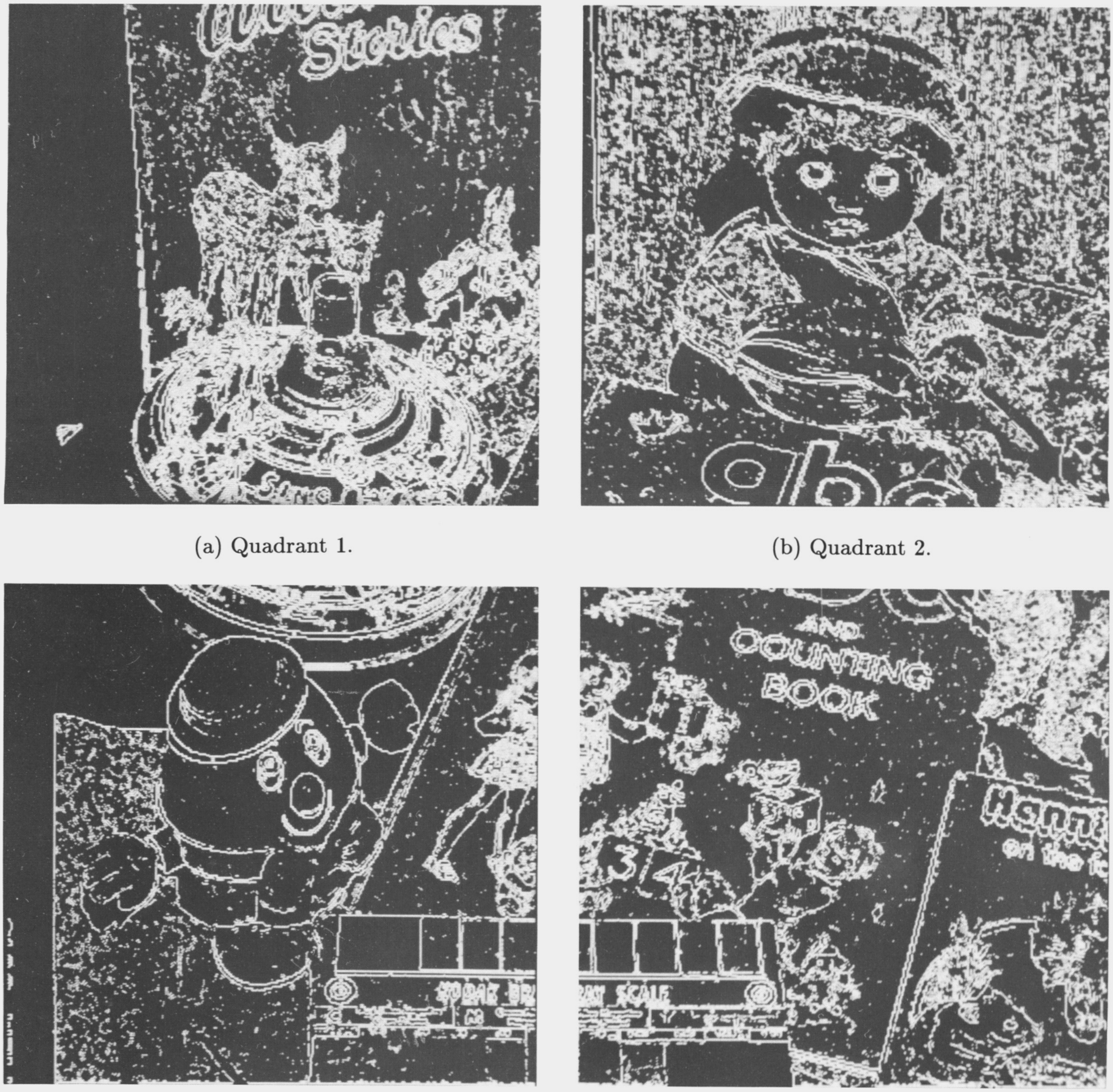
Figure 8. Smallest scale edge detection and representation process for 3 × 3 group of convolution samples.



(a) Toy scene.

(b) Mandrill.

Figure 9. Test images.



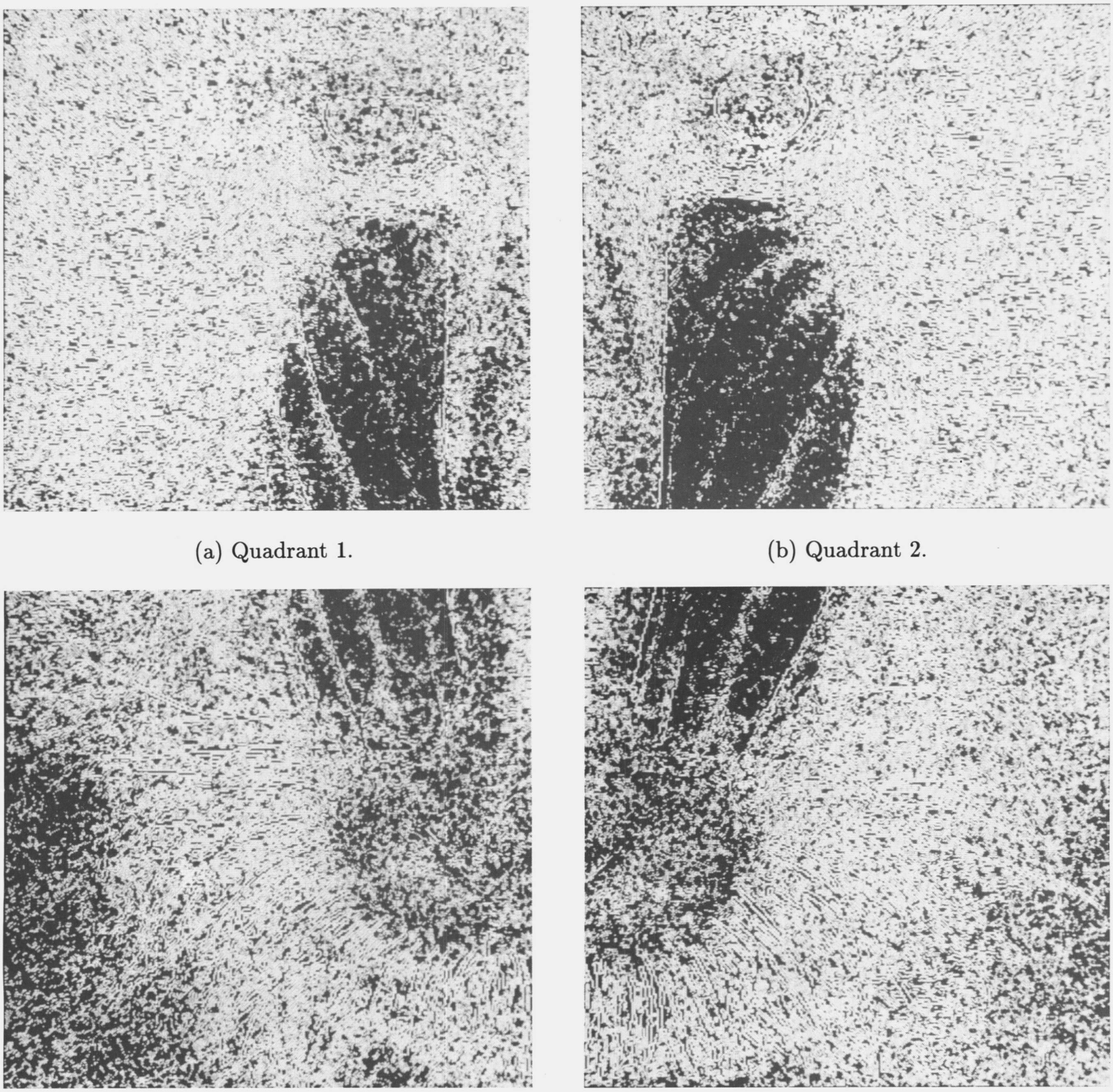
(a) Quadrant 1.

(b) Quadrant 2.

(c) Quadrant 3.

(d) Quadrant 4.

Figure 10. Smallest scale detected edge representation for toy scene.



(a) Quadrant 1.

(b) Quadrant 2.

(c) Quadrant 3.

(d) Quadrant 4.

Figure 11. Smallest scale detected edge representation for mandrill.

ORIGINAL PAGE
COLOR PHOTOGRAPH

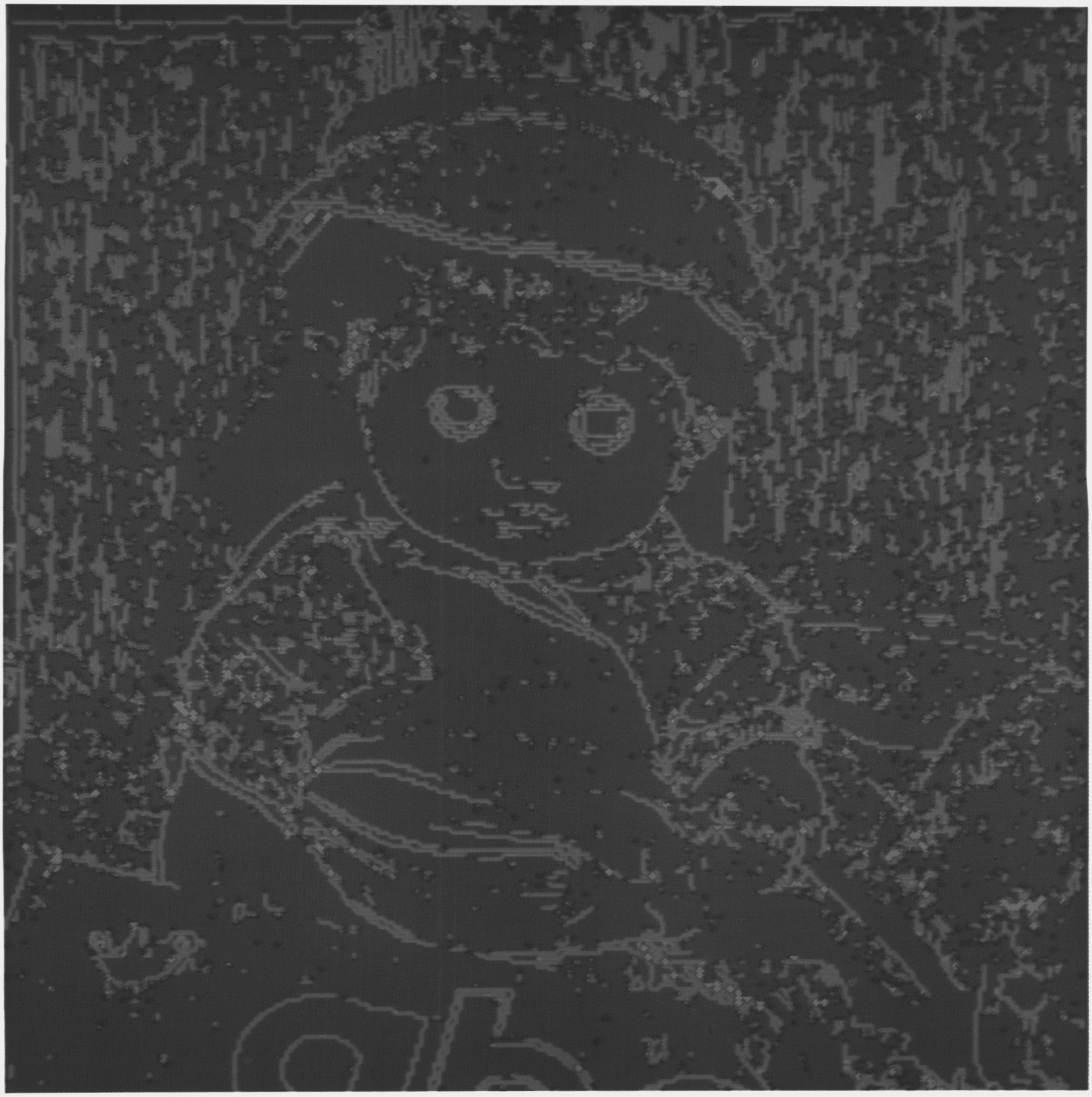


Figure 12. Smallest scale spatial primitive representation for quadrant 2 of toy scene.

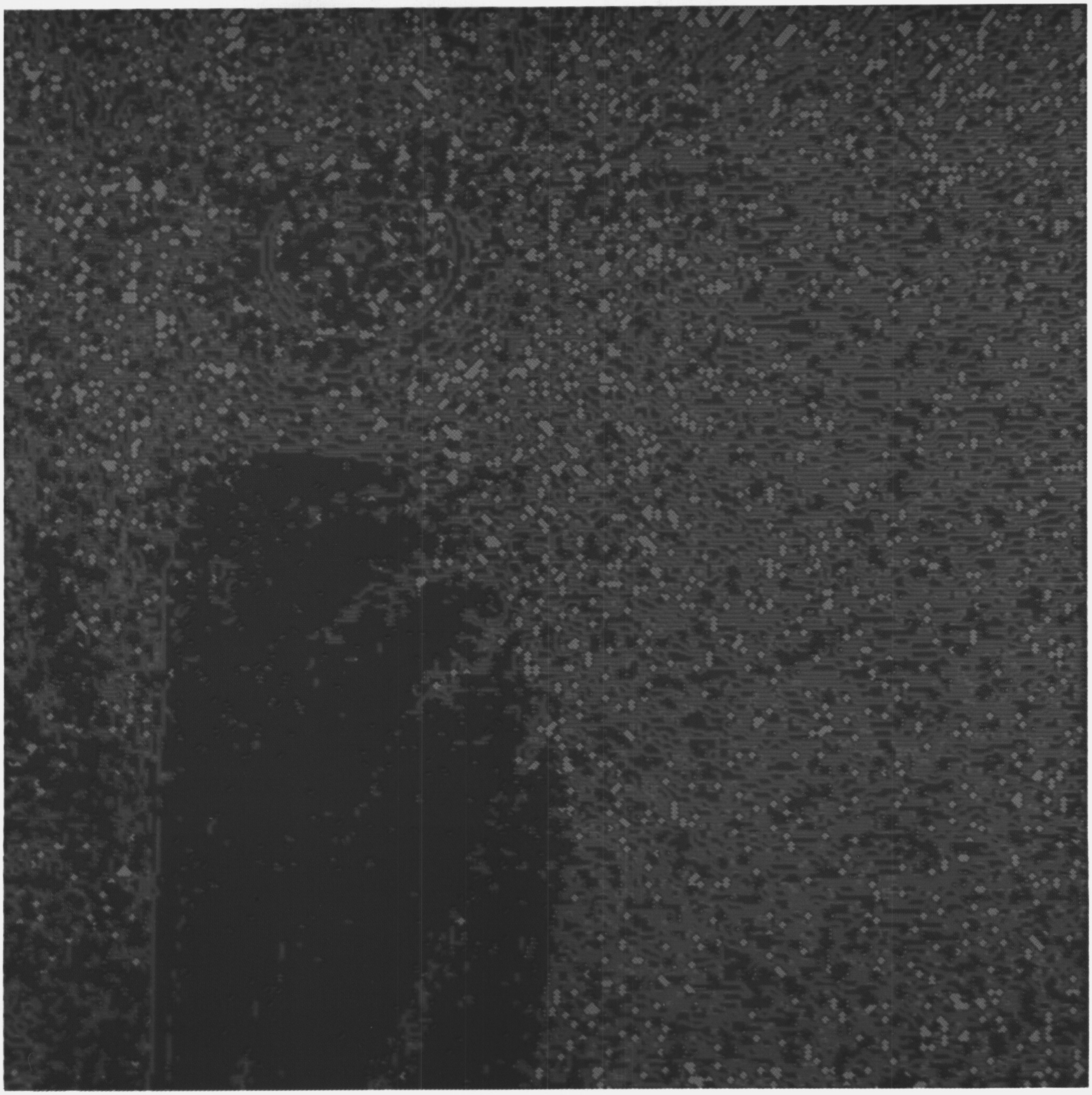
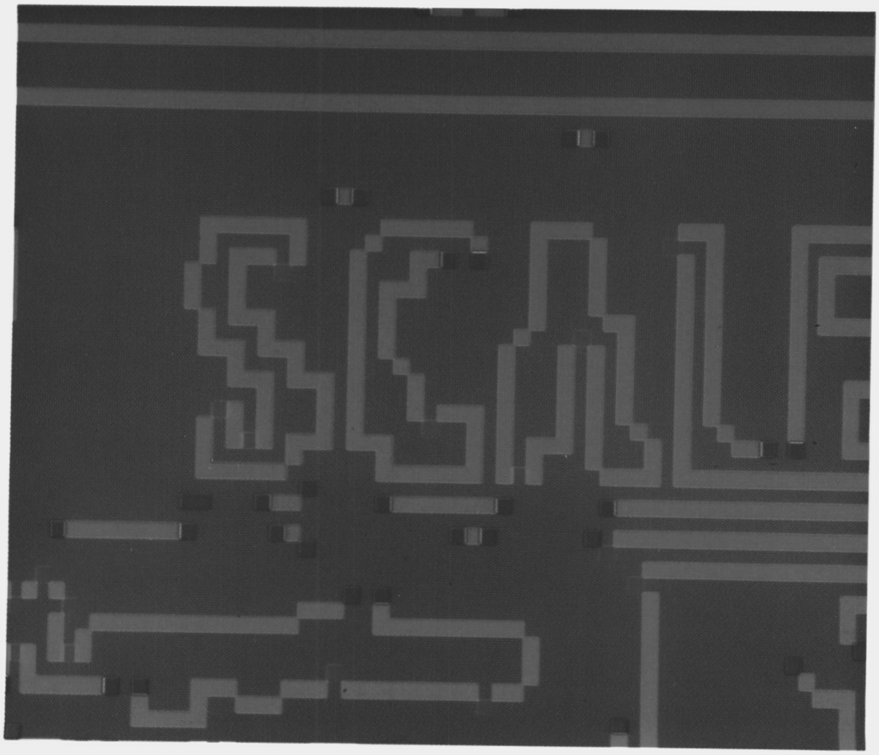
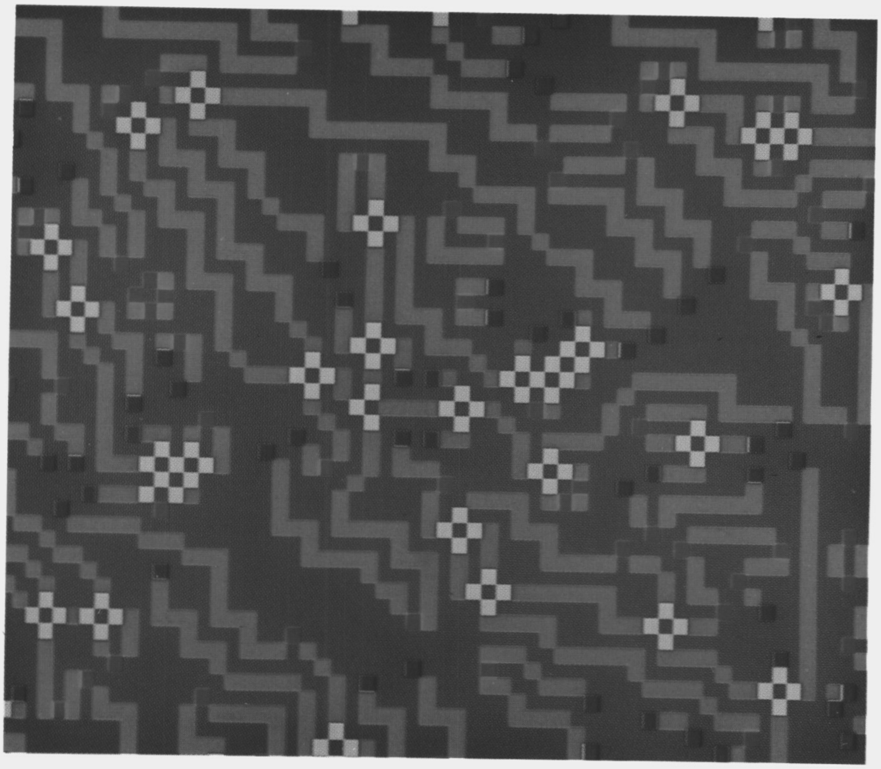


Figure 13. Smallest scale spatial primitive representation for quadrant 2 of mandrill.

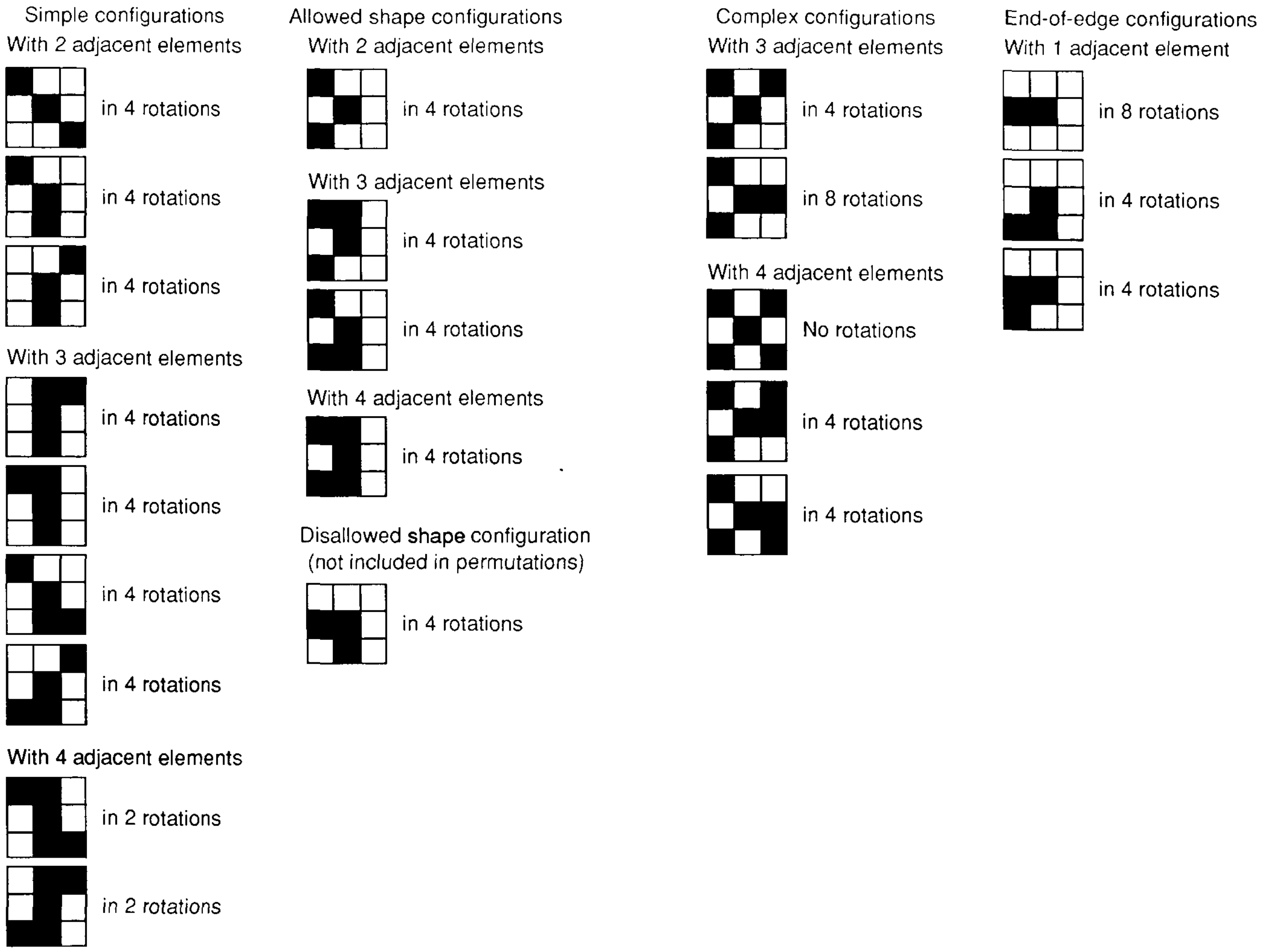


(a) Toy scene.



(b) Mandrill.

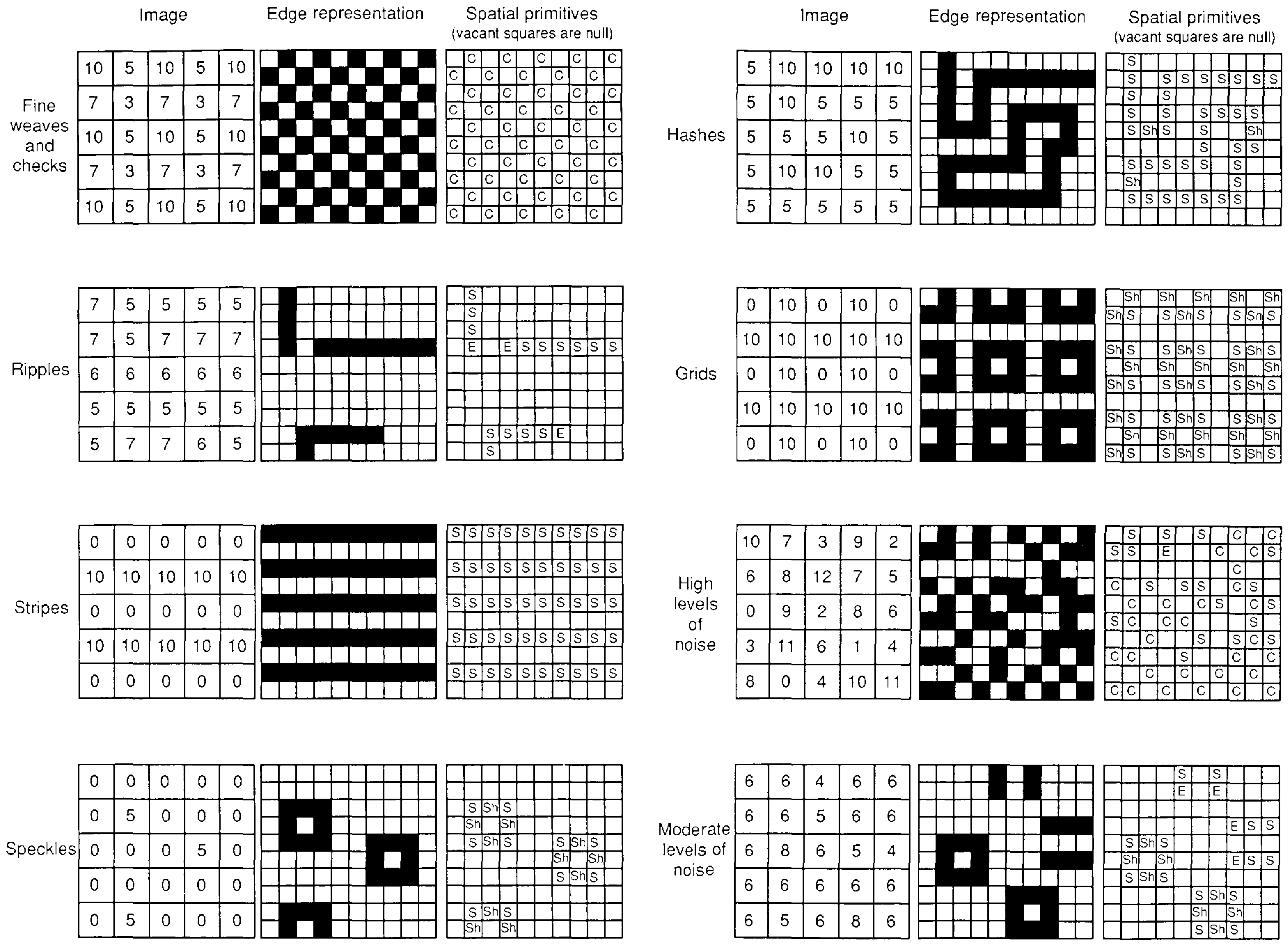
Figure 14. Details of smallest scale spatial primitive representation.



Edge permutations	24	16	21	16
Spatial primitive pattern permutations	2240	1600	3072	160

Total detected edge permutations = 77 out of 512 (2^9) possible
Total spatial primitive pattern permutations = 7072 out of 1 953 125 (5^9) possible

Figure 15. Exposition of detected 3 × 3 edge patterns and associated spatial primitive permutations (nonnull center elements).



Fine weaves and checks

10	5	10	5	10
7	3	7	3	7
10	5	10	5	10
7	3	7	3	7
10	5	10	5	10

Ripples

7	5	5	5	5
7	5	7	7	7
6	6	6	6	6
5	5	5	5	5
5	7	7	6	5

Stripes

0	0	0	0	0
10	10	10	10	10
0	0	0	0	0
10	10	10	10	10
0	0	0	0	0

Speckles

0	0	0	0	0
0	5	0	0	0
0	0	0	5	0
0	0	0	0	0
0	5	0	0	0

Hashes

5	10	10	10	10
5	10	5	5	5
5	5	5	10	5
5	10	10	5	5
5	5	5	5	5

Grids

0	10	0	10	0
10	10	10	10	10
0	10	0	10	0
10	10	10	10	10
0	10	0	10	0

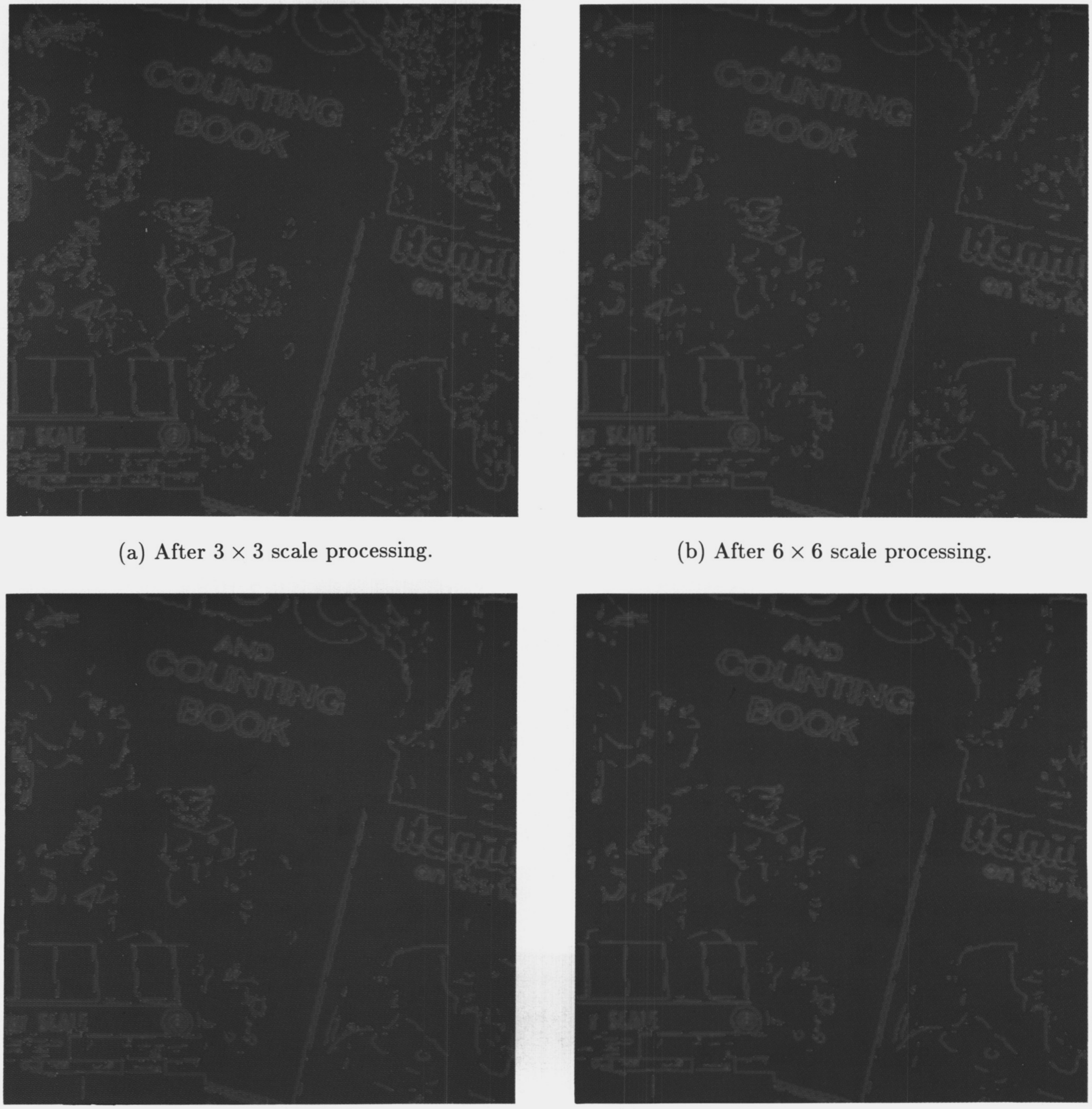
High levels of noise

10	7	3	9	2
6	8	12	7	5
0	9	2	8	6
3	11	6	1	4
8	0	4	10	11

Moderate levels of noise

6	6	4	6	6
6	6	5	6	6
6	8	6	5	4
6	6	6	6	6
6	5	6	8	6

Figure 16. Some classes of texture.



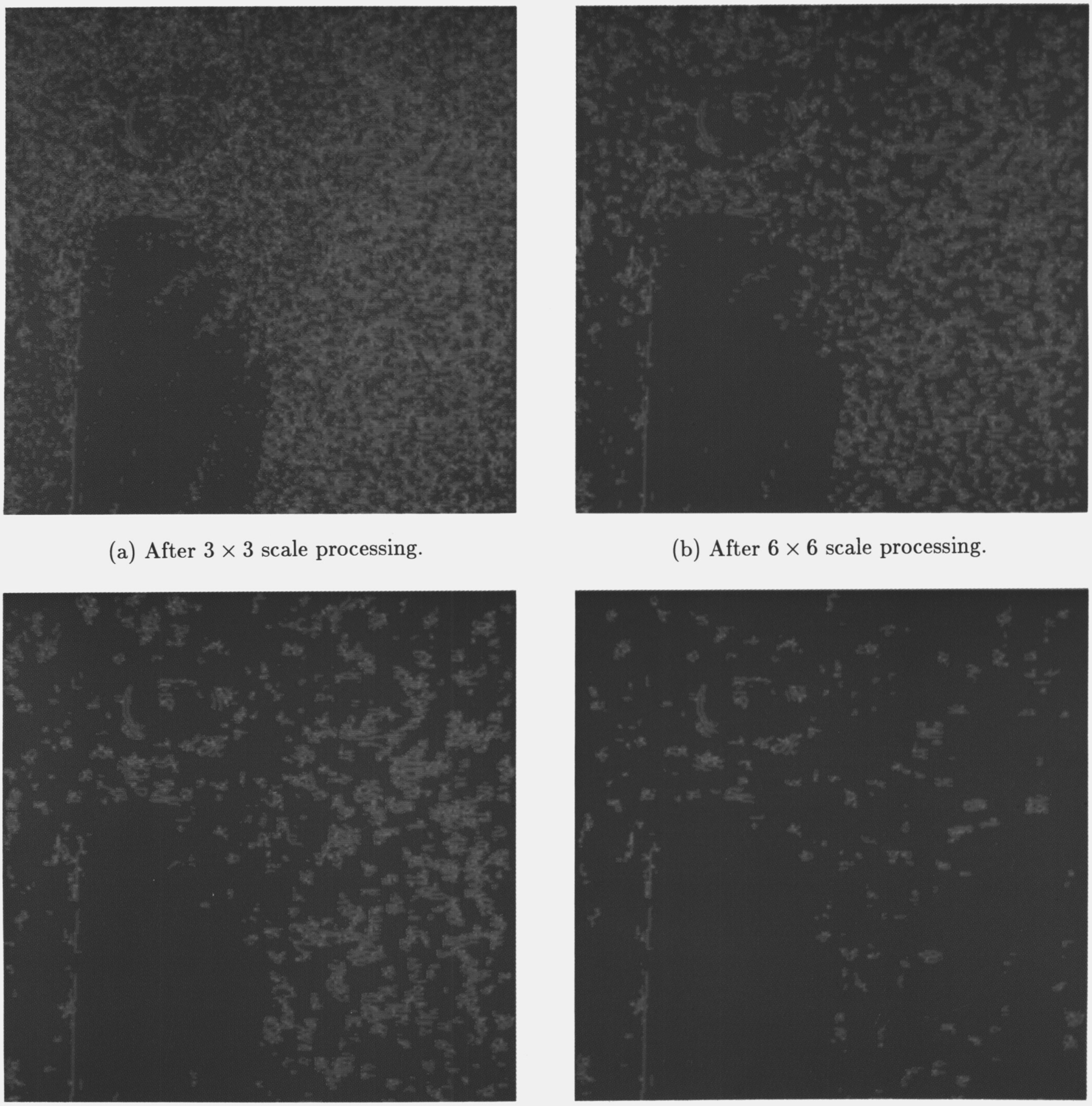
(a) After 3 × 3 scale processing.

(b) After 6 × 6 scale processing.

(c) After 12 × 12 scale processing.

(d) After 24 × 24 scale processing.

Figure 17. Sequence of multiscale contour operations for quadrant 4 of toy scene.



(a) After 3 × 3 scale processing.

(b) After 6 × 6 scale processing.

(c) After 12 × 12 scale processing.

(d) After 24 × 24 scale processing.

Figure 18. Sequence of multiscale contour operations for quadrant 2 of mandrill.

National Aeronautics and Space Administration

Report Documentation Page

1. Report No.
NASA TP-2838

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Spatial Vision Processes: From the Optical Image to the Symbolic Structures of Contour Information

5. Report Date
November 1988

6. Performing Organization Code

7. Author(s)
Daniel J. Jobson

8. Performing Organization Report No.
L-16479

9. Performing Organization Name and Address
NASA Langley Research Center
Hampton, VA 23665-5225

10. Work Unit No.
549-02-21-01

11. Contract or Grant No.

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Washington, DC 20546-0001

13. Type of Report and Period Covered
Technical Paper

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract
The significance of machine and natural vision is discussed together with the need for a general approach to image acquisition and processing aimed at recognition. An exploratory scheme is proposed which encompasses definition of spatial primitives, intrinsic image properties and sampling, two-dimensional edge detection at the smallest scale, construction of spatial primitives from edges, and isolation of contour information from textural information. Concepts drawn from or suggested by natural vision at both the perceptual and the physiological level are relied upon heavily to guide the development of the overall scheme. The scheme is intended to provide a larger context in which to place the emerging technology of detector-array focal-plane processors. The approach differs from many recent efforts at edge detection and image coding by emphasizing edge detection at the smallest scale as a foundation for multiscale symbolic processing while diminishing somewhat the importance of image convolutions with multiscale edge operators. Cursory treatments of information theory illustrate that the direct application of this theory to structural information in images could not be realized.

17. Key Words (Suggested by Authors(s))
Machine vision
Edge detection
Edge representation
Spatial primitive classification
Line-drawing information extraction
Natural-vision concepts
Imaging system
Texture removal

18. Distribution Statement
Unclassified–Unlimited

Subject Category 35/63

19. Security Classif.(of this report)
Unclassified

20. Security Classif.(of this page)
Unclassified

21. No. of Pages
29

22. Price
A03

NASA-Langley, 1988

For sale by the National Technical Information Service, Springfield, Virginia 22161-2171